

Crash Course in Statistics

Petchara Pattarakijwanich

SCPY208, 18 January 2020

เรื่องการเรียนรู้วิชานี้

- เวลาเรียน ทุกวันจันทร์ สิบโมงเช้าถึงเที่ยง
- เรียนออนไลน์ทั้งหมด Zoom Meeting ID: 327 782 3913
- การบ้าน 50% สอบกลางภาค 25% สอบปลายภาค 25%
- การบ้านทำส่งแบบไหนก็ได้ (ไม่บังคับ LaTeX)
- สอบเป็น take home (รายละเอียดค่อยประกาศ)
- ตัดเกรดอิงเกณฑ์ ตามปกติ
- ไม่บังคับเข้าเรียน (แต่ถ้าเข้าได้ก็ดี)
- ไม่บังคับให้เปิดกล้อง (แต่เปิดได้ก็ดี)
- วิดีโอและเลกเชอร์นี้จะอัปโหลดหลังคาบ

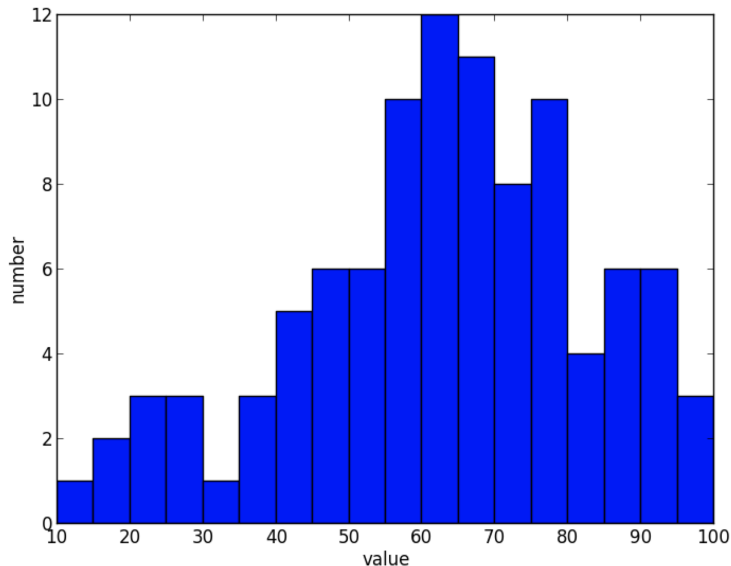
เครื่องมือทางสถิติที่น่าจะได้ใช้ รู้หายาๆ ไว้ก่อน เรียนอย่างละเอียดภายในเทอมนี้

- ❶ สถิติเพื่อการแสดงข้อมูล
- ❷ ความคลาดเคลื่อนของการวัด
- ❸ Error propagation
- ❹ การฟิตเส้นตรงกับข้อมูล

ตัวอย่าง คะแนนสอบของเด็กหนึ่งร้อยคนตามตารางด้านล่าง รายงานข้อมูลอย่างไร?

57	64	22	98	29	48	75	40	44	47
76	66	43	76	53	65	89	50	65	47
62	70	45	58	60	52	41	37	62	60
16	67	72	92	75	48	65	76	59	80
28	100	94	58	77	66	48	67	85	57
57	64	74	91	52	75	69	28	57	62
64	72	24	66	51	36	75	80	52	82
57	66	43	97	12	63	79	24	61	63
81	90	57	37	92	89	56	72	73	76
69	93	30	86	73	61	88	18	88	73

ตัวอย่าง คะแนนสอบของเด็กหนึ่งร้อยคนตามตารางด้านล่าง รายงานข้อมูลอย่างไร?



สองวิธีที่ทำได้

1 ราวๆ 100 ตัว

ข้อดี ครอบคลุม ครอบคลุม (information content สูงสุด)

ข้อเสีย เข้าใจยาก

2 ราวๆ ตัวเลข 2 ตัว

mean
SD

ข้อดี เข้าใจง่าย

ข้อเสีย ได้อย่างไร 2 ตัว อย่างไร?

สถิติ 50 ตัว

ข้อเสีย คือ (information มาก เข้าใจยาก)

↓ สถิติ

ข้อเสีย คือ (information น้อย เข้าใจง่าย)

Summary Statistics

Location

- Mean, Median, Mode
- Percentile, Quartile, Quintile, Decile
- Five-number summary (0, 25, 50, 75, 100)
- Seven-number summary (0, 10, 25, 50, 75, 90, 100)

Spread

- Standard Deviation, Variance
- Range, Interquartile Range

Shape

- Skewness
- Kurtosis

ตัวอย่าง วัดความสูงของนาย ก 10 ครั้งได้ดังนี้ รายงานอย่างไร? ฉลาดแค่ไหน?

179.87 173.16 173.42 171.78 177.60

168.10 180.23 172.72 175.96 174.25

$$\text{mean} = \frac{1}{10} \sum x_i \quad [\text{ตรวจสอบทราบค่า}]$$

mean เป็น estimator ของค่าเฉลี่ย

$$SD = \sqrt{\frac{1}{10-1} \sum (x_i - \text{mean})^2}$$

← ค่าความคลาดเคลื่อนของหน่วยวัด 1 ครั้ง

$$\text{ความคลาดเคลื่อนนี้ทราบ} \approx \frac{SD}{\sqrt{10}} \quad \leftarrow \text{ความคลาดเคลื่อนของ mean}$$

$$\text{รายงาน : ค่าเฉลี่ย} = \text{mean} \pm \frac{SD}{\sqrt{n}}$$

Sample mean และ sample SD คำนวณได้จาก

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

s เป็นความคลาดเคลื่อนของการวัด 1 ครั้ง

$\frac{s}{\sqrt{n}}$ เป็นความคลาดเคลื่อนของ $\bar{x}$

Central Limit Theorem
• x มาจากช่วงใดก็ได้
→ $\bar{x}$ มาจากแบบ Gaussian

รายงานผลดังนี้

$$\text{height} = \bar{x} \pm \frac{s}{\sqrt{n}}$$

ตัวอย่าง สมมติว่าวัดค่า x และ y ได้ดังนี้

$$x = \bar{x} \pm \sigma_x$$

$$y = \bar{y} \pm \sigma_y$$

อาจเป็น SD
หรือว่า $\frac{SD}{\sqrt{n}}$
แล้วแต่ context

ต้องการคำนวณค่า w ต่อไปนี้จาก x และ y ควรจะรายงานค่าอย่างไร?
คลาดเคลื่อนเท่าไร?

- $w_1 = x + y$
- $w_2 = xy$
- $w_3 = w(x, y)$ (กรณีทั่วไป ฟังก์ชันใดๆ)

$$w = x + y$$

$$\bar{w} = \bar{x} + \bar{y}$$

$$\sigma_w = \sqrt{\left(\frac{\partial w}{\partial x}\right)^2 \sigma_x^2 + \left(\frac{\partial w}{\partial y}\right)^2 \sigma_y^2}$$

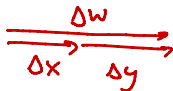
$$\sigma_w = \sqrt{\sigma_x^2 + \sigma_y^2}$$

$$x = \bar{x} \pm \Delta x$$

$$y = \bar{y} \pm \Delta y$$

$$w = x + y$$

$$\Delta w = \Delta x + \Delta y$$



$$\Delta w = \Delta x + \Delta y \quad ? \quad \text{หรือ} \quad \Delta w = \sqrt{(\Delta x)^2 + (\Delta y)^2}$$

$$w = xy$$

$$\bar{w} = \bar{x}\bar{y}$$

$$\sigma_w = \sqrt{\left(\frac{\partial w}{\partial x}\right)^2 \sigma_x^2 + \left(\frac{\partial w}{\partial y}\right)^2 \sigma_y^2}$$

$$\sigma_w = \sqrt{\bar{y}^2 \sigma_x^2 + \bar{x}^2 \sigma_y^2}$$

$$\frac{\sigma_w}{\bar{w}} = \sqrt{\left(\frac{\sigma_x}{\bar{x}}\right)^2 + \left(\frac{\sigma_y}{\bar{y}}\right)^2}$$

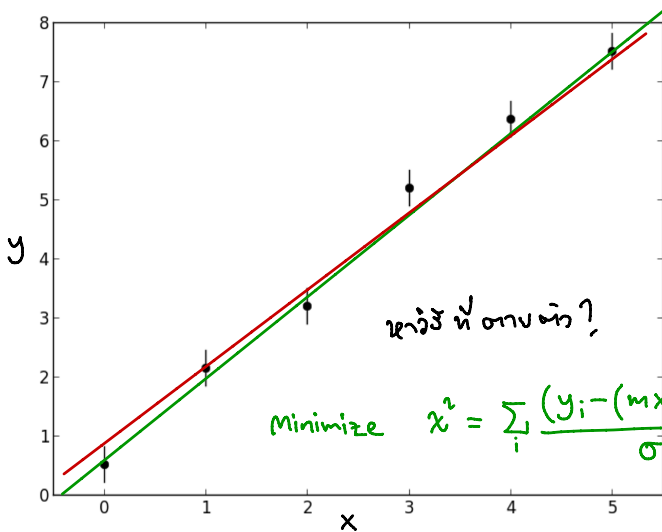
ตัวอย่าง สมมติว่ามีจุดข้อมูล (x, y, σ_y) ต่อไปนี้ พิตเส้นตรงยังไง?

x	y	σ_y
0	0.5	0.3
1	2.1	0.3
2	3.2	0.3
3	5.2	0.3
4	6.4	0.3
5	7.5	0.3

ตัวอย่าง สมมติว่ามีจุดข้อมูล (x, y, σ_y) ต่อไปนี้ พิตเส้นตรงยังไง?

$$y = mx + c$$

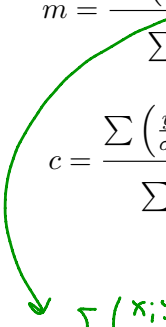
$m = ?$
 $c = ?$



พิตข้อมูลด้วย $y = mx + c$ หา m และ c ได้โดย

$$m = \frac{\sum \left(\frac{1}{\sigma_i^2} \right) \sum \left(\frac{x_i y_i}{\sigma_i^2} \right) - \sum \left(\frac{x_i}{\sigma_i^2} \right) \sum \left(\frac{y_i}{\sigma_i^2} \right)}{\sum \left(\frac{1}{\sigma_i^2} \right) \sum \left(\frac{x_i^2}{\sigma_i^2} \right) - \left(\sum \left(\frac{x_i}{\sigma_i^2} \right) \right)^2}$$

$$c = \frac{\sum \left(\frac{y_i}{\sigma_i^2} \right) \sum \left(\frac{x_i^2}{\sigma_i^2} \right) - \sum \left(\frac{x_i}{\sigma_i^2} \right) \sum \left(\frac{x_i y_i}{\sigma_i^2} \right)}{\sum \left(\frac{1}{\sigma_i^2} \right) \sum \left(\frac{x_i^2}{\sigma_i^2} \right) - \left(\sum \left(\frac{x_i}{\sigma_i^2} \right) \right)^2}$$


$$\sum \left(\frac{x_i y_i}{\sigma_i^2} \right) = \sum_{i=1}^n \left(\frac{x_i y_i}{\sigma_i^2} \right)$$

$$= \frac{x_1 y_1}{\sigma_1^2} + \frac{x_2 y_2}{\sigma_2^2} + \dots + \frac{x_n y_n}{\sigma_n^2}$$

ถ้า σ_i เท่ากันสำหรับทุกจุดจะลดรูปเหลือ

ตัวแปร ด้านล่างของสูตร

$$m = \frac{n \sum (x_i y_i) - \sum (x_i) \sum (y_i)}{n \sum (x_i^2) - (\sum (x_i))^2}$$

$$c = \frac{\sum (y_i) \sum (x_i^2) - \sum (x_i) \sum (x_i y_i)}{n \sum (x_i^2) - (\sum (x_i))^2}$$

x	y	x^2	xy	σ_y
0	0.5	0	0.0	0.3
1	2.1	1	2.1	0.3
2	3.2	4	6.4	0.3
3	5.2	9	15.6	0.3
4	6.4	16	25.6	0.3
5	7.5	25	37.5	0.3
15	24.9	55	87.2	

$$m = \frac{6 \times 87.2 - 15 \times 24.9}{6 \times 55 - (15)^2} = 1.43$$

$$c = \frac{24.9 \times 55 - 15 \times 87.2}{6 \times 55 - (15)^2} = 0.59$$

กฏของนิวตัน $y = ab^x$ (power-law)

$y = ae^{bx}$ (exponential)

$$\left. \begin{aligned} y = ab^x &\Rightarrow \ln y = \ln a + x \ln b \\ y = ae^{bx} &\Rightarrow \ln y = \ln a + bx \end{aligned} \right\} \begin{array}{l} \text{เส้นตรง} \\ (x, \ln y) \end{array}$$

$\Rightarrow$ ใช้วิธีแปลงเส้นตรงได้ บอกได้ว่า Error bar ของ $\ln y$

$$\sigma_{\ln y} = \frac{\partial \ln y}{\partial y} \sigma_y = \frac{1}{y} \sigma_y$$